ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/7966093

Development and Validation of a Brief Version of the Dyadic Adjustment Scale With a Nonparametric Item Analysis Model

Article *in* Psychological Assessment · April 2005

DOI: 10.1037/1040-3590.17.1.15 · Source: PubMed

CITATIONS
223

READS
1,841

3 authors, including:

Stéphane Sabourin
Laval University

150 PUBLICATIONS **3,109** CITATIONS

SEE PROFILE

Pierre Valois
Laval University

123 PUBLICATIONS **1,921** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

Evaluation of young athletes' dietary practices and those recommended by their coaches View project

Project

A cluster-analytic study of pornography use in a community sample View project

All content following this page was uploaded by Pierre Valois on 05 June 2014.

The user has requested enhancement of the downloaded file.

Psychological Assessment
2005, Vol. 17, No. 1, 15–27

Copyright 2005 by the American Psychological Association
1040-3590/05/$12.00 DOI: 10.1037/1040-3590.17.1.15

Development and Validation of a Brief Version of the Dyadic Adjustment Scale With a Nonparametric Item Analysis Model

Stéphane Sabourin and Pierre Valois
Université Laval

Yvan Lussier
Université du Québec à Trois-Rivières

The main purpose of the current research was to develop an abbreviated form of the Dyadic Adjustment Scale (DAS) with nonparametric item response theory. The authors conducted 5 studies, with a total participation of 8,256 married or cohabiting individuals. Results showed that the item characteristic curves behaved in a monotonically increasing function for all items of the brief 4-item version of the DAS (DAS-4). The DAS-4 proved to be informative at all levels of couple satisfaction. Compared with the 32-item version of the DAS (DAS-32), it was as effective in predicting couple dissolution and was significantly less contaminated by socially desirable responding. In addition, structural equation modeling demonstrated that the underlying latent construct measured by the DAS-4 was very stable over a 2-year period. This brief version has the advantage of being less time consuming and constitutes a promising alternative to the original DAS-32 for clinicians and researchers.

Despite considerable advances in the scientific study of couple processes and trajectories, progress in the measurement of marital quality has lagged behind. There are now numerous sophisticated conceptual and clinical models specifying the multiple contextual, transitory, and stable intrapersonal variables predicting marital adjustment as well as diverse relational outcomes. In most of these models (e.g., the vulnerability–stress–adaptation model, the interdependence model, and the individual differences model) and in several large-scale longitudinal studies of marriage (Gottman & Levenson, 2002; Kurdek, 1999; Lawrence & Bradbury, 2001), the central mediational role of couple satisfaction is recognized. However, the measurement of marital satisfaction continues to raise conceptual and methodological questions (Bradbury, Fincham, & Beach, 2000). Resolution of the controversy over multidimensional and unidimensional conceptualizations of couple satisfaction appears to be in favor of more restrictive single-factor approaches. However, this development has unfortunately encouraged the proliferation of nonstandardized, locally constructed, overly brief instruments mostly composed of redundant items (How happy are you with your relationship? How happy are you with your spouse?, etc.). Whereas these short measures generally have high face validity, there are multiple limits to this situation. For example, at the present time there are no adequate estimates of the prevalence of couple distress in the general community. Population studies, conducted with unstandardized self-report scales of couple satisfaction, generally show that more than 95% of married or cohabiting adults are very satisfied with their relationships (Glenn, 1998). Such estimates appear excessively high when confronted with social statistics showing that (a) couple problems are one of the chief complaints from persons consulting in individual psychotherapy (Maling, Gurtman, & Howard, 1995), (b) union dissolution rates are still at very high levels, ranging from 50% for first marriages to 70% for remarried partners (Statistics Canada, 2002; U.S. Bureau of the Census, 1999), and (c) the rates of coercive behaviors indicating the presence of physical, sexual, and psychological violence in close relationships (ranging from 10% to 25% for both physical and sexual violence and to 75% for psychological violence; for a recent review, see Archer, 2000), countering the optimistic portrayal of generalized couple bliss derived from national surveys of couple satisfaction.

This contradiction between survey results and social statistics can be explained by numerous methodological factors (the presence of sampling problems, the use of limited cross-sectional panels, the lack of information about specific couple trajectories, etc.). The use of nonstandardized measures could also be an important source of bias in results obtained. Most of these instruments were not constructed with classical or modern psychometric principles. There are no clear guidelines to determine which of these heterogeneous measures should become the main criterion from which cross-study comparisons and established norms could be developed. Thus, for both clinical and epidemiological purposes, there is a great need for brief standardized well-validated measures of couple distress. However, traditional measures such as the Marital Adjustment Test (MAT; Locke & Wallace, 1959) and the Dyadic Adjustment Scale (DAS; Spanier, 1976) are obviously too time consuming to be used in national surveys (even if they are still widely used for clinical and research purposes).

Shortened or abbreviated versions of these standardized and widely disseminated questionnaires provide an interesting avenue, which could be more fully explored before investing massive resources in the development of completely new instruments. The

---

Stéphane Sabourin, École de psychologie, Université Laval, Québec, Canada; Pierre Valois, Département des fondements et pratiques en éducation, Faculté des sciences de l'éducation, Université Laval; Yvan Lussier, Département de psychologie, Université du Québec à Trois-Rivières, Québec, Canada.

This research was supported in part by grants to Stéphane Sabourin and Yvan Lussier from the Fonds québécois de la recherche sur la société et la culture and from the Centre de recherche interdisciplinaire sur les problèmes conjugaux et les agressions sexuelles. We thank Hélène Paradis and Éric Frenette from the Research Unit on Children's Psychosocial Maladjustment for their assistance with the statistical analyses of the data.

Correspondence concerning this article should be addressed to Stéphane Sabourin, École de psychologie, Université Laval, Québec G1K 7P4, Canada. E-mail: stephane.sabourin@psy.ulaval.ca

DAS is certainly one of the questionnaires most widely used in research and clinical practice to determine the severity of couple problems, to decide whether couples will be offered preventive or therapeutic interventions, and to assess treatment progress. Even though there are diverse diagnostic cutoff points used for classification purposes, researchers use a large variety of total DAS scores, ranging from 92 to 107, to discriminate distressed from nondistressed couples (Baucom & Hoffman, 1986; Crane, Allgood, Larson, & Griffin, 1990; Eddy, Heyman, & Weiss, 1991). Because of its well-established psychometric properties, the DAS could serve as a good starting point to develop a brief cost-effective and valid measure of couple distress.

The DAS is not without its critics. Some argue that the multifactorial nature of the measure can lead to artificial overlapping between couple adjustment and interpersonal processes in marriage (Bradbury et al., 2000). Confounds between the assessment of relationship satisfaction and determinants of relationship satisfaction should certainly be avoided. However, it has also been clearly demonstrated that the primary factors of the DAS are almost perfectly related to a higher order couple adjustment factor and that the role of satisfaction in this hierarchical solution is crucial (Kurdek, 1992; Sabourin, Lussier, Laplante, & Wright, 1990).

To date, three independent groups of researchers have proposed brief versions of the DAS. These groups adopted different strategies to select the items of the short form they proposed: (a) a discriminant analysis: identifying items that maximally distinguish distressed from nondistressed couples (Sharpley & Cross, 1982; Sharpley & Rogers, 1984), (b) a mixed quantitative strategy: relying on multiple regression analyses and discriminant analyses to determine which scale of the DAS better predicts relationship quality and outcomes (Kurdek, 1992), and (c) a mixed two-step strategy: combining results from qualitative content analysis and confirmatory analysis (Busby, Christensen, Crane, & Larson, 1995). There are three frequently mentioned brief versions of the DAS. Sharpley and Rogers (1984) proposed a 7-item measure (DAS-7; Items 8, 10, 11, 25, 27, 28, 31), Kurdek (1992) suggested a 10-item questionnaire composed of items from the Satisfaction subscale of the DAS (DAS-10; Items 16 to 23 and Items 31 and 32) and Busby, Christensen, Crane, and Larson (1995) recommended a 14-item instrument (DAS-14; Items 3, 4, 6, 7, 12, 15, 16, 20, 21, 22, 24, 25, 27, and 28). Subsequent reliability and validity studies have shown that these three versions are consistent (Cronbach's alpha ranging from .76 to .96) and converge substantially with traditional full-scale couple adjustment measures such as the MAT, the DAS, and other measures of relationship quality variables (e.g., relationship beliefs, marital adaptability). Furthermore, both Sharpley and Rogers (1984) and Kurdek (1992), and more recently Hunsley, Best, Lefebvre, and Vito (2001), have demonstrated that their short DAS-based measures discriminated between married, separated, and divorced couples.

Whereas each of these short versions represents an interesting alternative to longer forms, they differ considerably in their composition and in the development process that was used to determine the final items. In fact, an examination of these short scales reveals three distinct problems. First, only two of the three short forms can truly be considered short enough to be used as a standardized scale in community epidemiological surveys. Comparable measures of psychological distress used in national surveys typically contain less than 10 items. For example, the Kessler scales (K 10 and K 6), used in annual government health surveys in the United States and Canada as well as in the World Health Organization World Mental Health Survey (Kessler et al., 2003), contain 10 and 6 items, respectively.

In the present study, our goal was to develop a short form that would be no longer than approximately 5 or 6 items. Second, the overlap in items selected for the short scales is surprisingly narrow, ranging from 1 item common to the DAS-7 and the DAS-10 to 4 items common to the DAS-10 and the DAS-14. This minimal overlap is partly the result of the use of different statistical techniques. We believe that the development of short screening scales should be based on item response theory (IRT), as are the new generation of mental health epidemiological surveys (Kessler et al., 2003). Third, the samples on which the short scales were constructed also vary to a large extent, ranging from 95 to 545 participants. These sample sizes, except from the one with 545 participants, are somewhat on the lower end of typically recommended psychometric guidelines for instrument development. In the present study, we used many large samples covering both community and clinical populations. With these limitations in mind, our general goal in the present study was to develop a short version of the DAS using modern test theory. One limitation of traditional psychometric approaches is that the estimation of the metric properties of a scale depend on the characteristics of the sample under study. Numerous psychometricians and statisticians (Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991; Santor & Ramsay, 1998; Thissen & Steinberg, 1997) have reported that classical test theory presents important shortcomings: (a) The standard error of measurement applies to all scores in a particular population; (b) unbiased estimates of item properties depend on having representative samples, which is rarely the case in psychology; (c) longer tests are more reliable than shorter tests; and (d) mixed item formats (which are frequent in the DAS) that use a Likert-type scale with 2-, 5-, 6-, and 7-point response formats lead to an unbalanced impact on a test's total score (Embretson & Reise, 2000).

For these reasons, psychometricians have sought alternative theories and models of mental measurement. Nunnally and Bernstein (1994) have already noted that IRT methods can be regarded as the central component of modern psychometrics and that although these techniques have been widely used with tests of skills, they should be used more often with psychosocial scales. In IRT, the standard error of measurement differs across but generalizes across populations, unbiased estimates of item properties may be obtained from unrepresentative samples, shorter tests can be more reliable than longer tests, and mixed item formats can yield optimal test scores.

A Nonparametric Item Response Analysis of the DAS

Recently, Ramsay (1995) developed a software program (TESTGRAF) based on nonparametric IRT and designed to estimate item and test characteristic curves of a test. The fundamental objective of this approach is to display the response curve that links the response to an option and θ, a latent variable. If the probability of endorsing an item or option changes as a function of, for instance, couple satisfaction, the item or option is effective. The aim of TESTGRAF is to estimate the response curve directly, thereby escaping the restrictions imposed by what can be achieved with a specific parametric family of curves. Of course, what is being fitted with TESTGRAF is the function itself, and not parameter estimates (Ramsay, 1997).

Thus, in the present study, TESTGRAF computes estimates of the probability of choosing an option as a function of the degree of couple satisfaction. This function curve is called the option characteristic curve. Once these option characteristic curves are estimated, on the basis of these curves, various other useful psychometric indices can also be estimated, including item characteristic curves, test reliability curves, and so forth.

This nonparametric item analysis approach has a number of important advantages over parametric or semiparametric approaches: (a) It provides useful information about how the choices offered vary as a function of the degree of endorsement of the attitude measured; (b) one is not constrained by characteristic curves, which can only exhibit features permitted by parametric models such as the two- or three-parameter logistic (2PL, 3PL) models, even when the data are sampled from a population in which the curves belong to the 2PL or 3PL family; and (c) the estimated characteristic curves have mean square errors as acceptable as those obtained by parametric models (see Ramsay, 1997, for a more detail theoretical presentation of the arguments for nonparametric models). A detailed description of this technique and the essentials of the algorithm have been published by Ramsay (2000).

Present Study

The main purpose of our research was to develop an abbreviated form of the DAS using IRT. We selected items that are generally effective (i.e., detect differences in couple distress severity and propose response options which are sensitive to changes in distress) and that provide maximum discrimination in the critical range of the trait, that is, around the diagnostic cutoffs of the DAS commonly used in the literature, scores ranging from 95 to 105. This interval was chosen because it contains most of the cutoff scores used by couple researchers using the DAS. In addition, this appears to represent a conservative estimate of the critical clinical zone in which diagnosis of couple distress is difficult to ascertain.

To attain this objective, we conducted five studies. In the first study, we conducted a nonparametric IRT analysis with TESTGRAF using 15 random splits of data from a large sample of participants. In the second study, we compared the performance of the abbreviated and original versions of the DAS in predicting union dissolution over a 3-year period. We hypothesized that the briefer version would be as robust as the longer one. In the third study, we examined whether social desirability was differentially related to couple adjustment in both the brief version and the 32-item version (DAS-32) of the DAS. In these three studies, we also compared our brief version with the other existing short forms of the DAS. The aim of the fourth study was to determine the presence of gender item biases. We examined differential item functioning to distinguish group mean differences from item bias with respect to how men and women endorse symptoms of distress on our self-report measures of couple satisfaction. Finally, the fifth study served to establish the temporal stability of our brief version of the DAS.

Study 1

Method

Participants and procedures. The sample consisted of 4,306 French-Canadian participants (2,153 couples) who had been married (59.29%) or cohabiting (40.71%). They were recruited on a voluntary basis through the media ($n = 2{,}526$), couple therapy services ($n = 424$), and infertility clinics ($n = 1{,}356$). Couples from the first group were informed that the study addressed various dimensions associated with close relationships. Couples from the second group were invited by their therapist on a voluntary basis to complete the DAS as part of a standard assessment protocol. Couples from the third group were invited by a member of the medical team to participate in a study dealing with the psychosocial repercussions of infertility. The mean age was 33.9 years ($SD = 8.8$) for the women and 35.9 years ($SD = 9.2$) for the men. The average number of years of cohabitation for the participants was 9.5 years. The average annual income was CAN\$35,760 (SD = CAN\$39,057) for the men and CAN\$20,528 (SD = CAN\$25,887) for the women. The mean number of children per couple was 1. All couples completed the DAS separately and were instructed to not discuss their responses with their partner.

Measure. The DAS, consisting of 32 items, is a comprehensive list of potential events that can occur in a couple's relationship. Respondents indicate the degree to which each event describes their couple adjustment during the past months using a Likert-type scale with 2-, 5-, 6-, and 7-point response formats. Standard scoring of the DAS yields a global distress score referred to as total marital adjustment, as well as scores for four subscales: Consensus (13 items), Satisfaction (10 items), Cohesion (5 items), and Affectional Expression (4 items). In the present study, we used the French-Canadian version of the DAS (Baillargeon, Dubois, & Marineau, 1986). The psychometric properties of this translation have been studied extensively. For example, the primary factors of the DAS are hierarchically related to a higher order factor (Sabourin et al., 1990), with higher order factor loadings ranging from .78 to .98. Reliability coefficients are always higher than .90. Likewise, as in studies with the original version of the DAS, validity coefficients are comparable and couple satisfaction scores from this French-Canadian version of the DAS have been related in similar ways to attachment styles (Lussier, Sabourin, & Turgeon, 1997), attributions (Lussier, Sabourin, & Wright, 1993), communication and/or problem solving behaviors (Bélanger, Sabourin, Laughrea, Dulude, & Wright, 1993), coping styles (Bouchard, Sabourin, Lussier, Wright, & Richer, 1998), personality (Bouchard, Lussier, & Sabourin, 1999), sex roles (Langis, Sabourin, Lussier, & Mathieu, 1994), and self-determination (Blais, Sabourin, Boucher, & Vallerand, 1990). For example, the pattern of relationship between measures of attributions concerning marital conflicts or negative spousal behaviors and DAS scores is consistent both in English (Bauserman, Arias, & Craighead, 1995) and in French (Lussier et al., 1993). The mean score of couple adjustment, computed at the individual level, for the general sample was 110.9 ($SD = 18.9$). When we used a cutoff score of lower than 100 on the DAS-32, 24.6% of the sample were distressed (547 women and 513 men). Although our sample was composed of couples, couple scores were not used. Data for the husbands and wives were included in the analyses as separate observations.

Results

Item selection. The item selection process adopted to derive a brief version of the DAS was based on a six-step approach. The first step was to randomly splice the general sample of 4,306 participants into equal subsamples of 2,153 participants. To minimize chance fluctuations, we repeated this process 15 times and thus obtained 30 different random subsamples of 2,153 participants. This randomization process was conducted with the restriction that only one member of each couple could be selected in a specific subsample. This constraint was applied because the partner scores on each item are nonindependent. The second step consisted of computing item characteristic curves for each of the 32 DAS items within each different subsample. In the third step, we computed the slope of these item characteristic curves around the diagnostic cutoff range of the DAS for each of the 32 items. We used an interval between 95 and 105 as the diagnostic cutoff range on the total score. The computation of these slopes was based on the following formula (Santor & Ramsay, 1998):

$$S_{iq} = \sum_{m} w_m P_{im}(\theta_q),$$

where S_{iq} is the value of the curve for item i at the qth evaluation point, w_m are the weights corresponding to each option, and $P_{im}(\theta_q)$ is the probability of endorsing option m of item i as a function of θ (latent variable or trait). Slope was calculated as follows:

$$\text{Slope} = \frac{S_{iq+1} - S_{iq-1}}{\theta_{q+1} - \theta_{q-1}}.$$

The fourth step consisted of ranking each item according to its slope; the item with the largest slope was ranked #1, the item with the second largest slope was ranked #2, and so on. This rank-ordering procedure was applied within each of the 30 randomly created subsamples. The fifth step was to determine, on the basis of the position of the item in each of the 30 trials, the number of times each item was ranked #1, #2 . . . #32. For example, as shown in Table 1, Item 31 was ranked first in all 30 random subsamples. Thus, in each trial, when we scrutinized the curves summarizing the relation between the scores on each item and the total scores on the DAS, this item was the one for which the slope at the critical 95–105 interval was always the steepest. This means that the item that was the most effective in discriminating individuals scoring between 95 and 105 on the DAS was Item 31, a general indicator of relationship happiness. Item 31 was also the most sensitive item at all levels of couple satisfaction (this will be discussed later). Note also, as shown in Table 1, the behavior of Item 19 (“Do you confide in your partner?”). This item was ranked second in 15 of 30 trials. It placed third in 14 trials, fourth in 1 trial, and fifth in 1 trial. Thus, the item measuring the capacity to confide in one’s partner was very sensitive to changes in couple satisfaction for people scoring between 95 and 105. Again, this item was effective across all levels of couple satisfaction. The last step was to determine the number of items for which the mean slope was higher than .40. In nonparametric IRT, the steeper a slope, the more discriminant the item is. However, there are no specific statistical criteria to determine whether a slope is significantly steeper than another. One must judge whether there is a specific demarcation in the slope distribution. In the present study, there was a clear demarcation between items in which the mean slope was higher or lower than .40. Using a mean slope of .40, we were also guaranteed that the number of items selected for a brief version of the DAS remained low.

Table 1
Position and Mean of Slopes for Each Item of the Dyadic Adjustment Scale With 15 Random Splits

	Position									
Item	1	2	3	4	5	6	7	*M*	*SD*	Median
1								.2415	.0242	.2435
2								.2272	.0155	.2316
3								.1151	.0175	.1136
4						1	4	.3859	.0250	.3807
5								.2280	.0161	.2268
6					1		2	.3724	.0247	.3683
7								.2148	.0201	.2167
8							2	.3783	.0236	.3819
9								.1991	.0205	.2007
10								.3165	.0180	.3202
11							1	.3713	.0179	.3717
12								.3422	.0164	.3422
13								.2727	.0235	.2755
14								.2984	.0165	.2911
15								.2090	.0187	.2130
16				4	17	7	2	.4251	.0225	.4270
17								.2155	.0194	.2159
18		14	14	2				.4773	.0244	.4798
19		15	14	1	1			.4777	.0309	.4728
20							1	.3694	.0187	.3705
21								.2590	.0203	.2612
22								.2987	.0195	.3002
23								.2726	.0185	.2765
24								.3204	.0167	.3225
25								.3501	.0253	.3513
26		1	2	21	2	3	1	.4391	.0224	.4383
27				2	7	9	4	.4137	.0279	.4166
28					3	10	13	.4080	.0196	.4098
29								.0644	.0077	.0646
30								.1477	.0113	.1467
31	30							.6574	.0320	.6634
32								.2060	.0125	.2061

Note. Position = number of times an item is selected in the first seven positions; M = mean of the 30 slopes for each item.

The results of these analyses are summarized in Table 1, which contains the means, standard deviations, and medians for the slopes obtained for all DAS items. The seven most frequently selected items, in order of importance, were Items 31, 19, 18, 26, 16, 27, and 28. The first three items were from the Satisfaction factor and the next one from the Cohesion factor. The last three most frequently selected items were from the Satisfaction and Cohesion factors. The mean of the slope was higher than .40 for all of these items. No items from the Consensus or Affective Expression factors showed up in this ranking. Five other items that showed up at least once in the first seven positions (Items 4, 6, 8, 11, and 20) were within an acceptable range but less efficient at the 95 to 105 range that we adopted to produce a short form of the DAS. Thus, Item 4 placed 5 times out of 30 in the sixth or seventh position, whereas Items 6, 8, 11, and 20 placed 3 times or less in the sixth or seventh position. The analyses also clearly showed that 20 of the 32 items never appeared in the first seven ranks (Items 1, 2, 3, 5, 7, 9, 10, 12, 13, 14, 15, 17, 21, 22, 23, 24, 25, 29, 30, and 32). The mean slope was lower than .35 for all of these items.

At this juncture, a close inspection of the results reveals that two different strategies could be adopted to form a brief version of the DAS. First, to maximize the theoretical purity of the scale, one could produce a short four-item version of the DAS by selecting items from the Satisfaction factor. The resulting DAS-4 would thus be formed of Items 16 (“How often do you discuss or have you considered divorce, separation, or terminating your relationship?”), 18 (“In general, how often do you think that things between you and your partner are going well?”), 19 (“Do you confide in your mate?”), and 31 (“Please circle the dot which best describes the degree of happiness, all things considered, of your relationship"). However, researchers looking for a strictly empirically based measure of couple distress that would maximally discriminate spouses at the 95 to 105 interval could use a 7-item version composed of Items 16, 18, 19, 26, 27, 28, and 31. Both the satisfaction and cohesion items from the original DAS would define this new scale. We selected the DAS-4 for theoretical reasons discussed previously: unidimensionality and minimization of the overlap between this measure of couple satisfaction and measures of other concomitant interpersonal processes in marriage. A comparison of these results with those observed for the three existing short versions of the DAS indicates that our DAS-4 shares one item with the DAS-7, four items with the DAS-10, and one item with the DAS-14. Among all items that are not shared with our seven most effective items, the mean slope ranged from .11 to .38. For items that were not part of our seven most effective items but which are retained in other short versions, none were ranked in the first four positions in any of the 30 random subsamples. The best of these nonselected items was Item 4, which ranked sixth, once out of 30, and seventh, four times out of 30.

Item characteristic curves. In the second step, we examined item characteristic curves to estimate the probability of choosing an option as a function of the degree of couple satisfaction. Results showed that the item characteristic curves behaved in a monotonically increasing function for all of the items of the DAS-4 (see Figure 1). Thus, expected item scores increased consistently across the entire spectrum of couple distress. The probability of endorsing higher options on these items increases as one's couple adjustment level becomes more positive. Consequently, these four items were effective not only in the critical interval we used as a selection criterion but also at all levels of distress. In Figure 1, for each item, vertical dashed lines in the item characteristic curves show the quantiles of the distribution of observed scale scores, and vertical bars on the curve in the graph on the right indicate estimated 95% confidence regions of the true curve. In contrast, analyses of item characteristic curves for rejected items[1] revealed that their scores did not climb as consistently as the total score increased—for example, Items 3, 17, and 32. In fact, for these items, even very distressed partners do not have the low expected item score they should obtain. Furthermore, these item characteristic curves are somewhat flattened (see Figure 1, Item 3). Item characteristic curves for the items in the other short DAS scales were also clearly less effective at a number of levels of couple distress. They can be divided into two groups. The first group is composed of flawed items for which the characteristic curve was generally flat (e.g., Items 3, 7, and 32); the second group consists of items that are more satisfactory, with acceptably steep curves where error variances at high levels of distress were large (e.g., Items 4, 6, and 8). The items in the second group could be included in a short version, but adding them would unnecessarily lengthen the scale.

Option characteristic curves. Option characteristic curves showed that for the four selected items, the probability of endorsing the option indicating the highest level of couple satisfaction (Option 5 for three of the four items; Option 6 for Item 31) increased as the latent trait, couple satisfaction, increased (option characteristic curves for the four items are plotted in Figure 1). For instance, for Item 31, "Please circle the dot which best describes the degree of happiness, all things considered, of your relationship," Option 6 (*Perfect*) is most frequently endorsed in the high range of couple satisfaction and the probability of this option being endorsed monotonically decreases as the level of couple satisfaction decreases. In fact, as expected, Option 6 begins to be chosen with a probability level higher than .20 only when the expected total score reaches 130. In a similar vein, the probability of being "Extremely unhappy" in one's relationship (Option 0) decreases steeply and almost reaches 0 when the total DAS score is higher than 80. The ranges at which other options used to rate couple happiness in Item 31 are easily identified; the curves change rapidly as couple happiness increases and options reach their peak in regions of the curve that correspond to their specific position in the couple happiness continuum. Thus, this item accurately reflects the individual differences in couple happiness at all levels of satisfaction. The behavior of Item 18 is very similar to the pattern of results observed for Item 31 and all options are effective. For Item 16 (thinking about separation or divorce), the probability of endorsing Option 0 (*All the time*) is problematic and is always lower than .15. It decreases to 0 when the total score on the DAS-32 reaches 80. Thus, the probability of even the most distressed spouses always thinking about separation or divorce is so low that this option should be discarded or reformulated. The same phenomenon is observed for Option 0 (*Never*) of Item 19 ("Do you confide in your partner?"). The probability of never confiding in your partner is so low that the effectiveness of this option can be questioned. The midpoint of the five options for this item (Option 3, *More often than not*) is also problematic because its characteristic curve is somewhat flat; its probability of being endorsed never goes higher than .20. This option can easily be confused with Option 4 (*Most of the time*). However, these items still fare much better in terms of endorsement probability with regard to their relationship with the latent trait, couple satisfaction, than the discarded items. In contrast, analyses of option characteristic curves for the rejected items mentioned previously (e.g., Item 3), showed that, for most of the options, the probability of endorsing the option indicating the highest level of couple satisfaction did not vary in accordance with their specific position in the couple happiness continuum.

Overall, the test information function indicated that the DAS-4 is informative at all levels of couple satisfaction. Reliability proved to be higher than .81 at all levels of couple distress. In fact, this lower bound was reached for cases in the first quartile of the sample. The reliability of the DAS-4 increased up to .92 for nondistressed participants. The traditional standardized alpha for the DAS-4 was .84. For purposes of comparison with the present data, the standardized alphas for the alternative brief versions used in previous studies were .85 for the DAS-7 (Sharpley & Rogers, 1984) and .88 for both the DAS-10 (Kurdek, 1992) and the DAS-14 (Busby et al., 1995). For the DAS-32, the alpha was .94. We tested to see whether the differences among these coefficients were significant, using the Charter and Feldt formula for dependent coefficients (Charter & Feldt, 1996). Because we used a very large sample, all the coefficients were significantly different, even the 1% difference between the DAS-4 and the DAS-7. However, the differences among the short versions of the DAS are not substantial and thus even a four-item version preserves good internal consistency.

---

[1] Option characteristic curves and expected item scores for all items of the DAS can be downloaded from the authors' Web site at http://www.uqtr.ca/lab_couple/Figures/

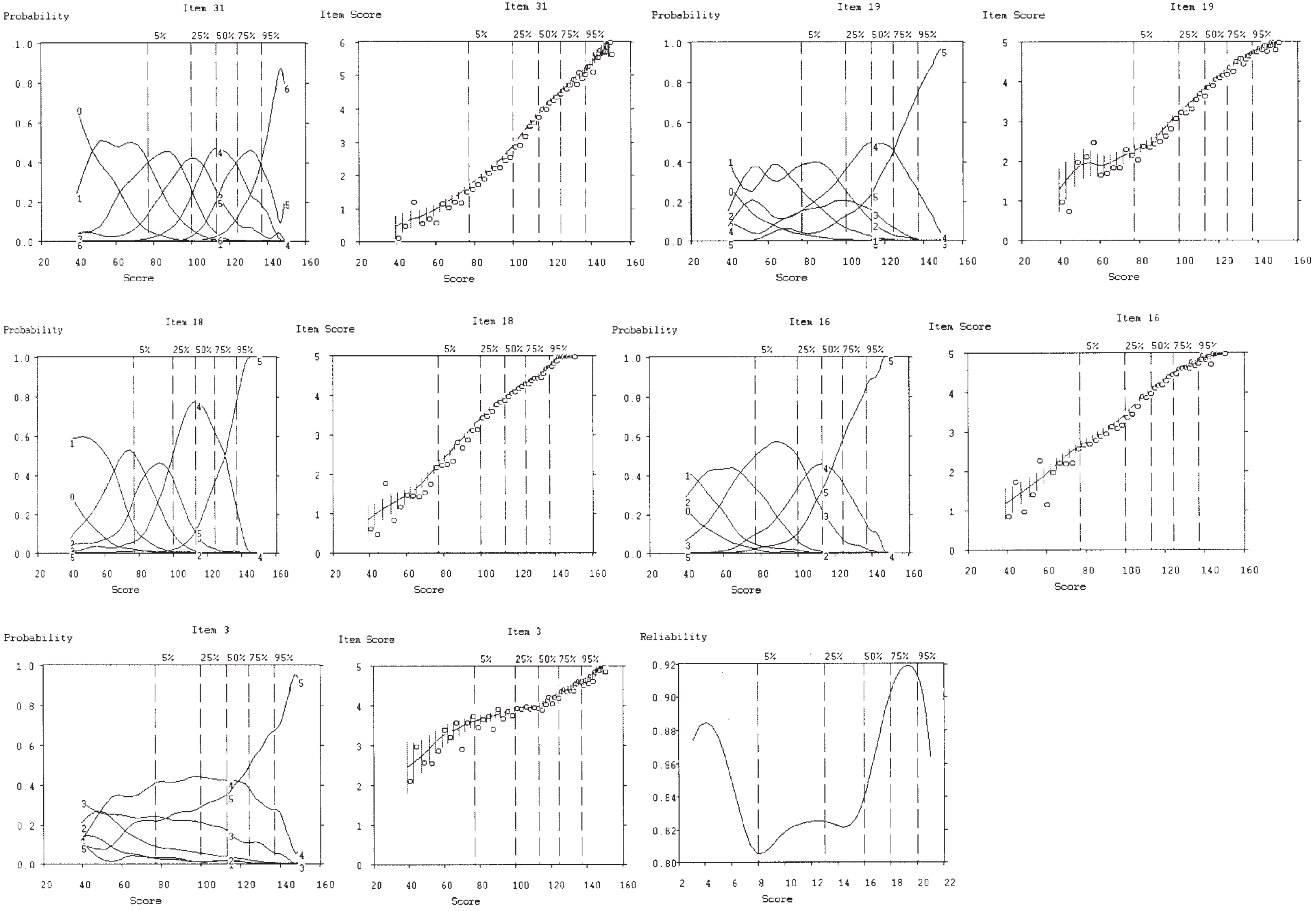

Figure 1. Option characteristic curves and expected item scores for Items 31, 19, 18, 16, and 3 and test information function (the right-most graph in the bottom row) for the four-item version of the Dyadic Adjustment Scale. The vertical dashed lines refer to various quantiles of the distribution of observed scores.

Table 2
Classification of Distressed and Nondistressed Participants According to Dyadic Adjustment Scale (DAS) Version

	Total sample					Sample excluding participants scoring in the gray zone (95–105)				
DAS version	Distressed (n = 1,060)	Nondistressed (n = 3,246)	Total (N = 4,306)	κ	CI	Distressed (n = 851)	Nondistressed (n = 2,923)	Total (N = 3,774)	κ	CI
DAS-4				.73	.70748, .75452				.83	.81244, .85556
n	829	3,053	3,882			730	2,828	3,558		
%	78.2	94.1	90.2			85.8	96.7	94.3		
DAS-7				.75	.72548, .77252				.87	.84740, .88660
n	861	3,044	3,905			754	2,846	3,600		
%	81.2	93.8	90.7			88.6	97.4	95.4		
DAS-10				.77	.75144, .79456				.87	.84840, .88760
n	867	3,080	3,947			752	2,850	3,602		
%	81.8	94.9	91.7			88.4	97.5	95.4		
DAS-14				.85	.83036, .86564				.95	.93524, .95876
n	976	3,080	4,056			825	2,878	3,703		
%	92.1	94.9	94.2			96.9	98.5	98.1		

Note. CI = confidence interval.

In the last step, we tested the relation between categorizations of distressed and nondistressed participants for the DAS-4, DAS-7, DAS-10, and DAS-14 compared with the DAS-32. Using a linear transformation, we computed a standardized z score for the DAS-32, adopting a cutoff of 100. This z score ($-.575$) was then applied to develop equivalent cutoff scores for all of the brief versions. For a person to be classified as a distressed spouse, the resulting cutoff scores were less than 13 for the DAS-4, less than 21 for the DAS-7, less than 34 for the DAS-10, and less than 45 for the DAS-14.[2]

Overall, as shown in Table 2, when we used the DAS-4, the reclassification rates of distressed and nondistressed participants were 78.2% and 94.1%, respectively; out of 4,306 partners, 3,882 were correctly reclassified, giving a global rate of 90.2%. However, a more detailed analysis of misclassified cases revealed that a large proportion of these errors could be observed in the range that corresponds to the 95 to 105 interval delineating ambiguous scores on the DAS-32. When scores in this range were excluded from the sample, reclassification rates increased to 85.8% for distressed and 96.7% for nondistressed spouses. For the DAS-7, DAS-10, and DAS-14, the classification rates were 90.7%, 91.7%, and 94.2%, respectively. The examination of kappas, and their confidence intervals, revealed that the only significant difference in these rates was that between the DAS-14 and all other brief versions. Thus, the DAS-14 was clearly more effective than the others because of its higher number of items. All of these kappas demonstrated more than substantial agreement above chance. Obviously, a cutoff score such as 13 for the DAS-4 must be viewed with caution. A simple change of one point of the cutoff score on a scale with a maximum score of 21 would lead to different reclassification rates, increasing or decreasing our ability to detect distressed and nondistressed participants. An approach based on a wider range of scores could also be effective and clinically meaningful. For example, scores between 12 and 14 on the DAS-4 correspond to scores in the critical clinical zone in which diagnosis of couple distress or satisfaction with the DAS-32 (95 to 105) is difficult to ascertain. It is in this problematic interval that classification rates were the lowest: 63.1% for the DAS-4, 60.1% for the DAS-7, 66.9% for the DAS-10, and 69.4% for the DAS-14. Thus, an alternative strategy for applying cutoff scores for the DAS-4 would consist of describing participants who score 12, 13, or 14 as borderline, those with scores under 12 as clinically distressed, and participants with scores over 14 as nondistressed.

For illustrative purposes, the DAS-4 total score for men and women in the couple therapy subsample are presented in Table 3. These clinical norms are compared with means and standard deviations obtained in both large scale studies conducted with representative samples of couples reported in Studies 4 and 5.

The results of this first study clearly revealed that the items that make up the DAS-4 show general effectiveness at all levels of couple distress and, more specifically, in the critical clinical zone in which diagnosis of couple distress is difficult to ascertain. Although alternative short forms such as the DAS-7, DAS-10, or DAS-14 have good reliability and maintain adequate classification rates of couple distress, many of their items appear to be less effective with respect to item characteristic curves and response options. Most of these items add little critical information relevant to the latent trait of couple satisfaction. The DAS-4 showed good reliability and an acceptable classification rate of couple distress as

---

[2] To determine whether these linear transformations were appropriate even if item characteristic curves imply a nonlinear function, using a parametric item response theory model (WINSTEPS; Linacre & Wright, 2000), we estimated theta scores (i.e., the score of each individual on the latent trait) for each participant on the DAS-32 and correlated the thetas with the standardized scores. This correlation reached .986. For the interval around the cutoff point (95–105), the relation between the thetas and the standardized scores is linear. Thus, standardized scores are satisfactory estimates of the relationship between raw scores on the DAS and the unobserved latent trait, couple satisfaction.

Table 3
Means, Standard Deviations, and Comparisons Between Men and Women for Clinical and Probabilistic Community Samples

	Women			Men				
Sample	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*df*
Clinical								
Study 1	212	10.18	3.48	212	10.41	3.38	0.90	211
Community								
Study 4	1,064	16.88	2.80	1,064	16.56	2.88	4.00***	1,063
Study 5								
Time 1	307	15.81	3.45	307	15.96	3.23	0.88	306
Time 2	144	16.10	3.18	144	16.01	3.06	0.39	143
Time 3	81	16.74	2.92	81	16.46	2.97	0.92	80

Note. Numbers of participants reported in this table are different from those of the original samples because we conducted paired *t* tests.
*** $p < .001$.

measured by the DAS-32. In addition, our version is short enough to be included in national or cross-national epidemiological studies of mental health and couple distress. For any short version of the DAS, there is a gray zone in which the decision to classify a participant as distressed or nondistressed is difficult to make. The DAS-4 cutoff score proposed in the present study will need to be scrutinized in future investigations. Finally, the DAS-4 shows minimal redundancy with closely related concepts in the field such as commitment, communication, and coping strategies. These are independent or dependent variables that are frequently used and correlated with dyadic adjustment. The use of an adjustment measure that covers the many facets of relationship quality and stability can create multicolinearity problems and artifactual relations between, for example, measures of couple adjustment (the Affectional Expression subscale of the DAS contains items which refer to sexual conflicts) and measures of sexuality.

There will be specific research or clinical situations in which it may be appropriate to use the full version of the DAS. For example, if theoretical purity is not an issue or if the goal is to target multidimensional facets or alternative domains of couple adjustment, a different combination of DAS items could be formed and studied. The DAS-32 covers a wide range of behaviors and situations representative of the multiple aspects of relationship quality (adjustment, divorce intentions, communication, household tasks, leisure activities, sexuality, etc.). Likewise, in couple therapy, the establishment of therapeutic objectives could be enriched by an examination of the more detailed content covered in the DAS-32.

Study 2

The aim of the second study was to examine the construct validity of the DAS-4. It has repeatedly been shown that couple satisfaction is a consistent predictor of relationship stability (see the review of Karney & Bradbury, 1995). However, it remains to be seen whether the briefer more theoretically focused DAS-4 that we developed preserves its predictive validity compared with the DAS-32 and with other existing short versions.

Method

Participants and procedures. The sample consisted of 175 French-Canadian couples (100 cohabiting and 75 married), recruited through various media (radio, TV, newspapers) and invited to take part in a study of the correlates of couple satisfaction. Participants were assessed twice over a 30-month period. Couple status was determined through a telephone interview at the end of the longitudinal study. Over a 30-month period, 56 of the couples had either separated or divorced (34 cohabiting and 22 married couples had either separated or divorced), whereas 119 were still living together. At their first visit to our laboratory, the cohabiting couples had been living together for an average of 3.7 (*SD* = 5.18) years, whereas the married couples had been living together for 12.6 (*SD* = 9.01) years. The cohabiting couples had an average 0.4 (*SD* = .95) children, whereas the married couples had 1.71 (*SD* = 1.33) children. The cohabiting and married women's mean ages were 33.12 (*SD* = 9.57) and 35.71 (*SD* = 8.51) years, respectively, and they had an average of 15.51 (*SD* = 3.42) and 16.16 (*SD* = 2.87) years of education and an annual income of CAN$21,604 (*SD* = CAN$15,128) and CAN$19,614 (*SD* = CAN$15,939). The cohabiting and married men's mean ages were 35.6 (*SD* = 10.69) and 37.53 (*SD* = 9.18) years, respectively, and they had an average of 15.61 (*SD* = 3.56) and 15.63 (*SD* = 3.54) years of formal education and earned an annual income of CAN$34,044 (*SD* = CAN$23,927) and CAN$35,627 (*SD* = CAN$20,138).

Measure. All participants completed the DAS-32 as part of an extended assessment battery.

Results

We conducted logistic regression analyses to determine the predictive power of the DAS-4, DAS-7, DAS-10, and DAS-14 with regard to couple dissolution. To circumvent the problem of nonindependence in couple data, we treated husbands and wives as separate observations, and we conducted distinct analyses for women and men. For both men and women, the results showed that the overall relation between each of these variables and couple dissolution was significant. For women, for the DAS-4, $\chi^2(1, N = 135) = 22.19$, $p < .01$, OR = .79, CI = .713, .884; for the DAS-7, $\chi^2(1, N = 135) = 18.45$, $p < .01$, OR = .88, CI = .824, .937; for the DAS-10, $\chi^2(1, N = 135) = 21.70$, $p < .01$, OR = .89, CI = .837, .938; for the DAS-14, $\chi^2(1, N = 135) = 22.65$, $p < .01$, OR = .91, CI = .868, .948; and for the DAS-32, $\chi^2(1, N = 135) = 23.27$, $p < .01$, OR = .96, CI = .947, .979. For men, for the DAS-4, $\chi^2(1, N = 135) = 22.65$, $p < .01$, OR = .79, CI = .705, .876; for the DAS-7, $\chi^2(1, N = 135) = 11.21$, $p < .01$, OR = .90, CI = .839, .959; for the DAS-10, $\chi^2(1, N = 135) = 20.23$, $p < .01$, OR = .87, CI = .820, .932; for the DAS-14, $\chi^2(1, N = 135) = 12.84$, $p < .01$, OR = .92, CI = .874, .965; and for the DAS-32, $\chi^2(1, N = 135) = 21.21$, $p < .01$, OR = .96, CI = .939, .977.

Thus, an examination of odds ratio estimates revealed that the probability of couple dissolution was best predicted by the DAS-4. For each one point increase on the DAS-4, the probability of couple dissolution decreased by 21% (1 – .79), whereas comparable estimates for the DAS-7 (12% for women and 10% for men), the DAS-10 (11% for women and 13% for men), the DAS-14 (9% for women and 8% for men), and the DAS-32 (4% for both women and men) were lower. There is therefore no loss in predictive validity when the DAS-4 is used with union stability as the criterion. In fact, there is a slight increase in predictive validity of the DAS-4 over the others.

Study 3

The goal of the third study was to determine whether socially desirable responding differentially affects couple satisfaction as measured by the original and shorter versions of the DAS. Social desirability is a significant bias that can lessen the validity of self-report scales in general. In couple research, social desirability is a particularly relevant issue because favorable self-presentation is a frequent goal in significant interpersonal relationships. In addition, the acknowledgment of couple distress can be psychologically threatening and thus researchers should strive to detect these biases.

Method

Participants and procedures. Participants were 678 French-speaking couples who were at their first appointment at a fertility clinic of a large metropolitan hospital. This sample was part of the larger sample used in Study 1. Couples participating in this study had to meet certain criteria: (a) They had never before received a complete diagnostic evaluation in a specialized clinic, (b) they were being treated for a problem of infertility, and (c) both partners had to be present. All participants had unsuccessfully attempted to conceive a child by natural means for a period of at least 1 year ($M = 38$ months, $SD = 26$ months) prior to intake. The average ages of the men and women were 32 and 30 years, respectively. The average annual income was CAN$34,782 for the men and CAN$20,108 for the women. The infertility problem was attributed to the female partner in 50% of the cases and to the male partner in 25% of the cases. In 13% of the cases, physicians attributed the problem to both partners; in 12% of the cases, the infertility was unexplained.

Measures. All participants completed the DAS-32 as part of an extended assessment battery. In addition, they completed the Balanced Inventory of Desirable Responding (Paulhus, 1984), a 40-item questionnaire that uses a 7-point scale to assess two factors defining social desirability: Self-Deception and Impression Management. The Self-Deception subscale (SDS) assesses the denial of thoughts about psychologically threatening but presumably universally true statements. Internal consistency coefficients for the SDS range from .70 to .80. Temporal stability coefficients over a 4-week period are satisfactory (.81). The construct validity of the SDS has been supported in numerous studies (Gur & Sackeim, 1979; Linden, Paulhus, & Dobson, 1986; Paulhus, 1982, 1984; Paulhus & Levitt, 1987; Roth & Ingram, 1985; Sackeim, 1983; Winters & Neale, 1985). In factor analyses, the SDS is at the center of a group of measures of defensive tendencies (Edward's Social Desirability Scale, Block's Ego Resiliency Scale, and Byrne's Repression-Sensitization Scale; Paulhus, 1984, 1986). The Impression Management subscale (IM) contains statements about socially desirable but infrequent behaviors. Alpha coefficients for the IM range from .72 to .75. The IM scores are significantly correlated with measures of psychological distress and physical health (Linden et al., 1986). In factor analyses, the IM appears at the center of a cluster of measures of impression management that includes the Minnesota Multiphasic Personality Inventory Lie Scale, the Wiggin's Social Desirability Scale, and the Positive Malingering Scale (Paulhus, 1984, 1986).

Results

Correlational analyses showed that both SDS and IM were significantly related to the DAS-32. For men and women, the correlations between self-deception and couple satisfaction were .27 ($p < .01$) and .26 ($p < .01$), respectively. For impression management, the coefficients were .29 ($p < .01$), and .31 ($p < .01$). For the DAS-4, again these correlations were significant, albeit weaker (men: SDS, $r = .17$, $p < .01$; IM, $r = .20$, $p < .01$; women: SDS, $r = .18$, $p < .01$; IM, $r = .25$, $p < .01$). We also tested whether the correlations with the DAS-32 and the DAS-4 were significantly different. The results showed that all correlations between self-deception, impression management, and couple satisfaction were significantly higher for the DAS-32 than for the DAS-4—men: SDS, $t(675) = 3.84$, $p < .01$; IM, $t(675) = 3.54$, $p < .01$; women: SDS, $t(675) = 3.81$, $p < .01$; IM, $t(675) = 2.70$, $p < .01$. We also computed these correlations for other existing short versions (DAS-7—men: SDS, $r = .19$, $p < .01$; IM, $r = .19$, $p < .01$; women: SDS, $r = .23$, $p < .01$; IM, $r = .23$, $p < .01$; DAS-10—men: SDS, $r = .22$, $p < .01$; IM, $r = .26$, $p < .01$; women: SDS, $r = .19$, $p < .01$; IM, $r = .27$, $p < .01$; DAS-14—men: SDS, $r = .23$, $p < .01$; IM, $r = .27$, $p < .01$; women: SDS, $r = .25$, $p < .01$; IM, $r = .29$, $p < .01$). Thus, for all short versions, the relations with social desirability are significant but low. In general, these coefficients were lower for the shorter versions than for the DAS-32. In fact, there were no significant differences between those observed for the DAS-4 and those observed for the DAS-7, whereas self-deception and impression management were more correlated with the DAS-10 than with the DAS-4 in men—SDS, $t(675) = 3.41$, $p < .01$; IM, $t(675) = 4.31$, $p < .01$—but not in women. Finally, self-deception and impression management were significantly more correlated with the DAS-14 than with the DAS-4—men: SDS, $t(675) = 2.03$, $p < .05$; IM, $t(675) = 2.28$, $p < .05$; women: SDS, $t(675) = 2.75$, $p < .01$—with the exception of impression management in women.

Study 4

The goal of the fourth study was to examine the differential item functioning of the DAS-4 to determine the presence of gender biases in selected items. Women generally report couple distress more frequently than men (Jacobson & Christensen, 1996). These results are most often interpreted as a substantive gender difference related to sociological or psychological factors. However, it is important to rule out the possibility of an artifactual difference associated with item bias. It is possible that the probability of endorsing an item is not equal for women and men even if they are equally distressed. Women and men might express their dissatisfaction in distinct ways. Nonparametric IRT can be used to distinguish qualitative differences in ways of expressing dissatisfaction from quantitative differences in severity of couple distress (Santor & Ramsay, 1998). It determines whether men and women with equivalent levels of couple distress endorse items in qualitatively different ways.

Method

Participants and procedures. The target population of the Quebec Longitudinal Study of Child Development (QLSCD) consisted of all babies, singleton births, born to mothers residing in Quebec over a 2-year period (1997 and 1998). Families were invited to participate in a study concerning factors associated with preschool child psychosocial adaptation. This representative sample was recruited through a stratified three-stage sampling design (Jetté & Des Groseilliers, 2000). The final sample was composed of 2,223 households, and the response rate for the first observation period of the longitudinal study (i.e., for 5-month-old infants) was 75.8%. Couple distress was assessed during the third observation period of the QLSCD when the children were 2½ years of age. From 1998 to 2000, 1,985 households provided data for the three observation periods, for a longitudinal response rate of 93.7%. Our specific sample included the biological parents of the targeted child who were still living together and still residing in Quebec. The final sample consisted of 1,586 mothers and 1,382 fathers.

Measure. Only items from the DAS-4 were administered in the present study. Data for the husbands and wives were included in the following analyses as separate observations.

Results

Expected item scores as a function of expected total couple satisfaction scores were plotted comparing women and men for each item of the DAS-4. These results revealed that expected item scores do not vary for women and men; in fact, for the four items, men's and women's item characteristic curves overlap to a considerable extent. Furthermore, in all four cases, the composite differential item functioning, a summary statistic for the amount of bias in each item, were all lower than .005. Thus, in the DAS-4, there were no qualitative differences between men and women in how they express couple dissatisfaction. However, paired *t* tests revealed significant quantitative gender differences on the total score of the DAS-4 as well as on all four items. Overall, women (M = 4.23, CI = 4.193, 4.266) were more satisfied than men (M = 4.15, CI = 4.112, 4.187), $F(1, 1352) = 20.96$, $p < .01$. However, η^2 for this significant effect was low at .015. More specifically, women were less likely than men to have considered separation or divorce (M = 4.52 for women, CI = 4.49, 4.57; M = 4.48 for men, CI = 4.41, 4.523), $t(1352) = 2.31$, $p < .05$, $\eta^2 = .02$. They reported more often than men that things were going well between them and their partner (M = 4.17 for women, CI = 4.134, 4.212; M = 4.08 for men, CI = 4.047, 4.123), $t(1351) = 4.28$, $p < .01$, $\eta^2 = .01$). Women, compared with men, reported that they confided more often in their partner (M = 4.25 for women, CI = 4.204, 4.303; M = 3.92 for men, CI = 3.855, 3.975), $t(1353) = 9.95$, $p < .01$, $\eta^2 = .07$, but they expressed less couple happiness than men (M = 3.93 for women, CI = 3.859, 3.996; M = 4.09 for men, CI = 4.027, 4.166), $t(1064) = -4.54$, $p < .01$, $\eta^2 = .02$. For each of the four comparisons, an examination of confidence intervals and η^2 values revealed that these significant differences were of the same magnitude as those observed for the total score of the DAS-4, that is, not substantial. Finally, because these data were not normally distributed, we also used a nonparametric method of comparison, namely, the Wilcoxon signed-ranks test. All differences remained significant. Means and standard deviations for men and women are presented in Table 3.

Study 5

The goal of the last study was to determine the temporal stability of the DAS-4 over a 2-year period. Past studies have shown that the DAS-32 is relatively stable over short-term periods (e.g., .87 over a 3-week period, Carey, Spector, Lantinga, & Krauss, 1993; .96 over an 11-week period, Stein, Girodo, & Dotzenroth, 1982). However, over a 4-year period (Kurdek, 1992), test–retest correlations were much lower (.48). It remains to be seen whether a much shorter version of the DAS is stable over both short- and long-term intervals.

Method

Participants and procedures. This representative random sample was made up of 316 French-Canadian heterosexual couples who had been married (n = 202) or cohabiting (n = 114) for at least 6 months and were older than 18 years. Spouses were asked to participate in a study concerning factors associated with couple relationships. The couples had been living together for an average of 13 years. Participants had an average of 1.64 children (n = 464; 73%). The average age of both men and women was 39 years, with 14 years of education. The average annual income of the men was CAN$40,000, whereas the average annual income of the women was CAN$20,000. Couples were recruited by way of a home survey. The sample was compiled with telephone numbers from across the Province of Quebec in Canada. Overall, 63.2% of couples returned their questionnaires. The margin of error for this study was 5.5, with a confidence level of 95%. As part of a longitudinal study, the couples were contacted on two other occasions over a 2-year period, once every 12 months. The retention rates for the second and third period were 46.7% (n = 295 participants) and 27.5% (n = 174 participants), respectively. We conducted structural equation analyses with participants who completed the measure at each of the three observation periods.

Measure. Only items from the DAS-4 were administered in the present study. Alpha was .91.

Results

We examined the temporal stability of the DAS-4 with structural equation modeling. The model specified significant relationships between each of the four indicators and the latent variable (the measurement model) as well as unidirectional relations between the latent variables from Time 1 to Time 3 (the structural model). The model was tested with PROC CALIS (SAS Institute, 1999) and allowed for significant correlated errors between related indicators over the three-time period. For example, the error associated with the first item from the DAS-4 at Time 1 was allowed to correlate with the error for the first item of the DAS-4 at Times 2 and 3, and so forth. The inclusion of these errors is standard in longitudinal models.

This model was tested to determine the invariance of the results across women and men. We adopted a two-step strategy. First, the preceding measurement and structural models were estimated separately for women and men. The fit indices for women (goodness-of-fit index [GFI] = .91; root-mean-square error of approximation [RMSEA] = .064; Bentler's comparative fit index [CFI] = .983), $\chi^2(39, N = 86) = 52.67$, $p > .05$, and men (GFI = .899; RMSEA = .084; Bentler's CFI = .963), $\chi^2(39, N = 86) = 61.17$, $p < .01$, were satisfactory. Thus, the models were reproduced across sex. Second, a series of equality constraints were added to the model, specifying that standardized estimates should not vary for men and women for either the measurement or the structural model. The fit indices of the invariant model were high (GFI = .89; RMSEA = .054; Bentler's CFI = .969). The chi-square to degree of freedom ratio was lower than 2, which is another indication of a good fit between the model and the data, $\chi^2(90, N = 172) = 133.21$, $p < .01$. For the structural model, the correlations between couple satisfaction at each point in time were equivalent for men and women. The correlation between couple satisfaction at Time 1 and couple satisfaction at Time 2 were significant and reached .87 for men and .83 for women. Couple satisfaction at Time 1 was significantly associated with couple satisfaction at Time 3 for women (.45) but not for men (.28). Finally, couple satisfaction at Time 2 was significantly related with couple satisfaction at Time 3: The correlation between the latent traits was .63 for men and .49 for women. Thus, the temporal stability of couple satisfaction from Time 1 to Time 2 and from Time 2 to Time 3 was quite high. These stability estimates are generally invariant across sex. The only exception to this pattern of invariance was the relation between DAS-4 scores from Time 1 to Time 3 for men. In this case, when couple satisfaction at Time 2 is entered into the structural equation, it acts as a mediator and the relationship between couple satisfaction at Time 1 and Time 3 loses its significance.

For the measurement model, the strength of the relation between the latent variable, couple satisfaction, and each indicator was significant. Results for the measurement model revealed that within each observation period a unidimensional factor structure was reproduced. The relation between the latent variable and each indicator was strong. At Time 1, factor loadings ranged from .57 to .81; at Time 2, factor loadings varied from .60 and .90; and at Time 3, factor loadings ranged from .58 to .86. The R^2 for each variable observed ranges from .32 to .80, and measurement errors in this model are therefore small. The only difference between men and women was noted for Item 18 ("Things are going well between my partner and I") at Time 1. The factor loadings for this item differed according to gender (women: $B = .634$, $R^2 = .402$; men: $B = .459$, $R^2 = .211$), $\chi^2(1, N = 154) = 4.0$, $p < .05$. Thus, it appears that this indicator is less strongly related to couple satisfaction for men than it is for women. Only at Time 1 does it explain couple satisfaction less effectively for men than for women. Means and standard deviations for men and women are presented in Table 3.

Discussion

The main goal of these five studies was to examine item performance in the DAS-32, to produce an effective short version of this widely used scale, and to determine its psychometric properties with nonparametric IRT. This new short version was also compared with other brief versions of the DAS-32. Results from item characteristic curves showed that seven items provided maximal discrimination at the critical range distinguishing distressed from nondistressed partners. Four of these seven items, all coming from the Satisfaction scale of the DAS, were selected to form the DAS-4. This selection preserves the theoretical consistency of the scale and its unidimensionality, as recommended in recent reviews of conceptual and methodological literature (Bradbury et al., 2000), and provides a brief standardized measure of couple satisfaction that can be used in community surveys as well as in clinical studies, thus facilitating between-study comparisons. By adopting a large interval ranging from 95 to 105 on the original DAS, the selected items were highly efficient in a clinically ambiguous zone, a factor that should enhance the clinical utility of the DAS-4. It is also important to note that these items also showed general effectiveness at all levels of couple distress. Furthermore, the high number of random splits produced with the total sample enhanced the validity of the item selection process.

The DAS-4 can still be used as a continuous measure of couple satisfaction and can also serve to categorize participants into distressed and nondistressed groups. The cutoff point, which corresponds to the classical threshold of 100 proposed by many authors, would now be at 13. This cutoff point results in a low rate of misclassification errors. Future studies are needed to determine the sources and consequences of these discrepancies with more precision. However, a critical test of this clinical threshold and determination of the classification properties of the DAS-4 should be conducted in studies based on both self-report scales and structured diagnostic interviews for couple distress (e.g., Heyman, Feldbau-Kohn, Ehrensaft, Langhinrichsen-Rohling, & O'Leary, 2001). In sum, the DAS-4 is a very brief measure globally targeting the content domain of couple satisfaction. It can be useful for screening purposes either in large scale studies in which it is not possible to administer the full DAS or in clinical contexts in which the main objective is not to assess couple satisfaction or adjustment per se but to evaluate clinical syndromes that can show comorbidity with marital problems.

The nonparametric item response model used in the present study led to the selection of four items with excellent precision at all levels of the couple satisfaction continuum. The predictive validity of this brief scale of couple satisfaction was supported in a 3-year longitudinal study of couple dissolution. In addition, the temporal stability of the latent construct measured by the DAS-4 was quite high over a 2-year period. Socially desirable responding was significantly less related to couple satisfaction in the DAS-4 than in the DAS-32. This could represent an improvement explained by the fact that we selected less threatening items that did not evoke self- or other-deceptive mechanisms. However, this could also be an artifact related to smaller variance in the DAS-4 than in the DAS-32. Difference in the lengths of the two measures could also account for these lower correlations. The use of representative samples of couples in two of our five studies provided strong community norms for the DAS-4. This will facilitate between-study comparisons. It will also serve as a criterion against which scores from more specific samples of partners—parents with sick children, married versus cohabiting couples, couples of other sexual orientations, couples in which one of the partners has medical or psychosocial difficulties—could be contrasted. Finally, the DAS-4 is composed of items that appear to be nonbiased with respect to gender. This will enable researchers to examine true quantitative differences in couple satisfaction between men and women; these sex variations cannot be attributed to distinct ways of understanding the items. In the present study, women were slightly more satisfied on each of the DAS-4 items as well as on the total score. However, these gender differences should not be overinterpreted given that they are quite small and most likely are explained by the large size of our sample.

The comparisons between the DAS-4 and all other existing brief versions of the DAS revealed that our shorter version is not only equal to the other measures with regard to its psychometric properties, but when applying IRT, our four items are also better indicators of the latent trait, couple satisfaction, as indexed by the DAS-32. This is especially important because content analyses of the other short versions of the DAS indicated little item overlap. If there had been a set of items common to the DAS-7, DAS-10, and DAS-14, this could have formed a consistent basis on which to develop an even shorter version of the DAS. The lack of convergence was thus another good reason to conduct the present set of studies. In addition, certain items from the other brief versions of the DAS that we rejected on empirical grounds showed wide theoretical and content heterogeneity, as they came from different subscales of the DAS that measure not only relationship quality but also psychological processes (e.g., communication, sexuality, etc.) associated with the concept of couple quality. There have been many criticisms of the conceptual and methodological ambiguities associated with these confounded omnibus measures of couple satisfaction (Bradbury et al., 2000).

There are some limitations to the present studies. First, despite its advantage as a screening instrument, the DAS-4 should not be considered as a replacement for the original questionnaire. The decision to produce a very short empirically based unidimensional form of the DAS could be seen by proponents of a multidimensional conception of dyadic adjustment as an arbitrary choice. Even if we applied a theoretical criterion to select, among the seven most effective items, the ones that would eventually form the DAS-4, these results would need to be reexamined from a theoretical perspective. For example, what is the relationship between satisfaction and adjustment? When spouses express dissatisfaction on the DAS-4, does it mean that they are distressed on all other facets of couple adjustment and vice versa? Traditionally, couple adjustment has been described as a more general inclusive concept than couple satisfaction. If couple adjustment includes, in addition to satisfaction, other meaningful specific facets, the content validity of our DAS-4 could be limited. Second, even if the cross-cultural pattern of relations between couple satisfaction and affective, cognitive, and behavioral measures of marital functioning appears quite stable, it may be important to replicate our results with English-speaking participants with a similar item analysis approach. Such replication efforts would also provide the much needed additional data in order to continue the examination of all of the other short forms derived from the DAS-32. As stated more than a decade ago, from a psychometric viewpoint the DAS-32 can still not be described as an overresearched measure (Sabourin et al., 1990). The same can be said for all short versions of the DAS. Furthermore, the existing short forms should not be banned or necessarily replaced by the DAS-4. There are still no definitive criteria to establish whether measures developed from IRT are theoretically and empirically superior to instruments developed with classical test theory. Third, an examination of option characteristic curves showed that some options in Items 16 (thinking about separation or divorce) and 19 ("Do you confide in your partner?") were problematic. In both cases, the Option 0 was almost never indicated at all levels of couple satisfaction. Some adjustments will therefore be necessary; for example, this option could be eliminated or reformulated (e.g., replacing "never" with "practically never") and the effectiveness of these modifications will have to be empirically tested. For now, and before such results are available, we recommend that clinicians and researchers use the DAS-4 as is. Fourth, some caution should be exercised when using DAS-4 scores to categorize participants into distressed or nondistressed groups. The DAS-4 is a self-report instrument and its convergence with other diagnostic measures should be examined. Structured clinical interviews and observational measures will help to put in context the merits and limits of self-report data. Finally, other studies will be needed to determine the DAS-4 sensitivity to therapeutic change.

References

Archer, J. (2000). Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin, 126,* 651–680.

Baillargeon, J., Dubois, G., & Marineau, R. (1986). Traduction française de l'échelle d'ajustement dyadique (A French translation of the Dyadic Adjustment Scale). *Revue Canadienne des sciences du comportement, 18,* 25–34.

Baucom, D. H., & Hoffman, J. A. (1986). The effectiveness of marital therapy: Current status and application to the clinical setting. In N. S. Jacobson & A. Gurman (Eds.), *Clinical handbook of marital therapy* (pp. 597–620). New York: Guilford Press.

Bauserman, S. A. K., Arias, I., & Craighead, W. E. (1995). Marital attributions in spouses of depressed patients. *Journal of Psychopathology and Behavioural Assessment, 17,* 231–249.

Bélanger, C., Sabourin, S., Laughrea, K., Dulude, D., & Wright, J. (1993). Macroscopic marital interaction coding systems: Are they interchangeable? *Behaviour Research and Therapy, 31,* 789–795.

Blais, M., Sabourin, S., Boucher, C., & Vallerand, R. J. (1990). Toward a motivational model of couple happiness. *Journal of Personality and Social Psychology, 59,* 1021–1031.

Bouchard, G., Lussier, Y., & Sabourin, S. (1999). Personality and marital adjustment: Utility of the five-factor model of personality. *Journal of Marriage and the Family, 61,* 651–660.

Bouchard, G., Sabourin, S., Lussier, Y., Wright, J., & Richer, C. (1998). Predictive validity of coping strategies on marital satisfaction: Cross-sectional and longitudinal evidences. *Journal of Family Psychology, 12,* 112–131.

Bradbury, T. N., Fincham, F. D., & Beach, S. R. H. (2000). Research on the nature and determinants of marital satisfaction: A decade in review. *Journal of Marriage and the Family, 62,* 964–980.

Busby, D. M., Christensen, C., Crane, D. R., & Larson, J. H. (1995). A revision of the Dyadic Adjustment Scale for use with distressed and nondistressed couples: Construct hierarchy and multidimensional scales. *Journal of Marital and Family Therapy, 21,* 289–308.

Carey, M. P., Spector, H. P., Lantinga, L. J., & Krauss, D. J. (1993). Reliability of Dyadic Adjustment Scale. *Psychological Assessment, 5,* 238–240.

Charter, R. A., & Feldt, L. S. (1996). Testing the equality of two alpha coefficients. *Perceptual and Motor Skills, 82,* 763–768.

Crane, D. R., Allgood, S. M., Larson, J. H., & Griffin, W. (1990). Assessing marital quality with distressed and nondistressed couples: A comparison and equivalency table for three frequently used measures. *Journal of Marriage and the Family, 52,* 87–93.

Eddy, J. M., Heyman, R. E., & Weiss, R. L. (1991). An empirical evaluation of the Dyadic Adjustment Scale: Exploring the differences between marital "satisfaction" and "adjustment." *Behavioral Assessment, 13,* 199–220.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Glenn, N. D. (1998). The course of marital success and failure in five American 10-year marriage cohorts. *Journal of Marriage and the Family, 60,* 569–576.

Gottman, J. M., & Levenson, R. W. (2002). A two-factor model for predicting when a couple will divorce: Exploratory analyses using 14-year longitudinal data. *Family Process, 41,* 83–96.

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology, 37,* 147–169.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* London: Sage.

Heyman, R. E., Feldbau-Kohn, S. R., Ehrensaft, M. K., Langhinrichsen-Rohling, J., & O'Leary, K. D. (2001). Can questionnaire reports correctly classify relationship distress and partner physical abuse? *Journal of Family Psychology, 15,* 334–346.

Hunsley, J., Best, M., Lefebvre, M., & Vito, D. (2001). The seven-item short form of the Dyadic Adjustment Scale: Further evidence for construct validity. *American Journal of Family Therapy, 29,* 325–335.

Jacobson, N. S., & Christensen, A. (1996). *Acceptance and change in couple therapy: A therapist's guide to transforming relationships.* New York: Norton.

Jetté, M., & Des Groseilliers, L. (2000). Survey description and methodology. In *Longitudinal study of child development in Québec (ELDEQ 1998–2002)* (Vol. 1, No. 1). Québec, Canada: Institut de la statistique du Québec.

Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, method, and research. *Psychological Bulletin, 118,* 3–34.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., et al. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60,* 184–189.

Kurdek, L. A. (1992). Dimensionality of the Dyadic Adjustment Scale: Evidence from heterosexual and homosexual couples. *Journal of Family Psychology, 6,* 22–35.

Kurdek, L. A. (1999). The nature and predictors of the trajectory of change in marital quality for husbands and wives over the first 10 years of marriage. *Developmental Psychology, 35,* 1283–1296.

Langis, J., Sabourin, S., Lussier, Y., & Mathieu, M. (1994). Masculinity, femininity, and dyadic adjustment: An examination of theoretical models. *Journal of Personality, 62,* 393–414.

Lawrence, E., & Bradbury, T. N. (2001). Physical aggression and marital dysfunction: A longitudinal analysis. *Journal of Family Psychology, 15,* 135–154.

Linacre, J. M., & Wright, B. D. (2000). WINSTEPS: Multiple-choice, rating scale and partial credit Rasch analysis [Computer Software]. Chicago: MESA Press.

Linden, W., Paulhus, D. L., & Dobson, K. S. (1986). Effects of response style on the report of psychological and somatic distress. *Journal of Consulting and Clinical Psychology, 54,* 309–313.

Locke, H. J., & Wallace, K. M. (1959). Short marital adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living, 22,* 251–255.

Lussier, Y., Sabourin, S., & Turgeon, C. (1997). The role of attachment and coping in marital relationships. *Journal of Social and Personal Relationships, 14,* 777–791.

Lussier, Y., Sabourin, S., & Wright, J. (1993). On causality, responsibility and blame in marriage: Validity of the entailment model. *Journal of Family Psychology, 7,* 1–11.

Maling, M. S., Gurtman, M. B., & Howard, K. (1995). The response of interpersonal problems to varying doses of psychotherapy. *Psychotherapy Research, 5,* 63–75.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory.* New York: McGraw-Hill.

Paulhus, D. L. (1982). Individual differences, self-presentation, and cognitive dissonance: Their concurrent operation in forced compliance. *Journal of Personality and Social Psychology, 43,* 838–852.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609.

Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires* (pp. 143–165). New York: Springer-Verlag.

Paulhus, D. L., & Levitt, K. (1987). Socially desirable responding triggered by affect: Automatic egotism? *Journal of Personality and Social Psychology, 52,* 245–259.

Ramsay, J. O. (1995). *TESTGRAF: A program for the graphical analysis of multiple choice test and questionnaire data.* Unpublished manuscript, McGill University, Montreal, Quebec, Canada.

Ramsay, J. O. (1997). A functional approach to modeling test data. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 381–394). New York: Springer.

Ramsay, J. O. (2000). *TestGraf: A program for the graphical analysis of multiple choice test and questionnaire data.* Unpublished manuscript, McGill University, Montreal, Quebec, Canada.

Roth, D. L., & Ingram, E. R. (1985). Factors in the self-deception questionnaire: Association with depression. *Journal of Personality and Social Psychology, 47,* 243–251.

Sabourin, S., Lussier, Y., Laplante, B., & Wright, J. (1990). Unidimensional and multidimensional models of dyadic adjustment: A hierarchical reconciliation. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2,* 333–337.

Sackeim, H. A. (1983). Self-deception, self-esteem, and depression: The adaptive value of lying to oneself. In J. Masling (Ed.), *Empirical studies of psychoanalytic theories.* (Vol. 1, pp. 101–157). Hillsdale, NJ: Erlbaum.

Santor, D. A., & Ramsay, J. O. (1998). Progress in the technology of measurement: Applications of item response theory. *Psychological Assessment, 10,* 345–359.

SAS Institute. (1999). *SAS/STAT user's guide, version 6* (4th ed.). Carey, NC: Author.

Sharpley, C. F., & Cross, D. G. (1982). A psychometric evaluation of the Spanier Dyadic Adjustment Scale. *Journal of Marriage and the Family, 44,* 739–741.

Sharpley, C. F., & Rogers, H. J. (1984). Preliminary validation of the abbreviated Spanier Dyadic Adjustment Scale: Some psychometric data regarding a screening test of marital adjustment. *Educational and Psychological Measurement, 44,* 1045–1049.

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family, 38,* 15–28.

Statistics Canada. (2002). *Changing couple life in Canada.* General social survey-Cycle 15 (Catalogue No. 85–576- XIE). Ottawa, Ontario, Canada: Author.

Stein, S. J., Girodo, M., & Dotzenroth, S. (1982). The interrelationship and reliability of a multilevel behavior-based assessment package for distressed couples. *Journal of Behavioral Assessment, 4,* 343–360.

Thissen, D., & Steinberg, L. (1997). A response model for multiple choice items. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 381–394). New York: Springer.

U.S. Bureau of the Census. (1999). *Unmarried-couples households, by presence of children: 1960 to present, AD-2.* Retrieved June 13, 2000, from http://www.census.gov/population/socdemo/ms-la/tabad-2.txt

Winters, K. C., & Neale, J. M. (1985). Mania and low self-esteem. *Journal of Abnormal Psychology, 94,* 282–290.

Received March 5, 2003
Revision received September 17, 2004
Accepted September 21, 2004 ■

View publication stats